NBER WORKING PAPER SERIES

# ESTIMATING THE IMPACT OF CROP DIVERSITY ON AGRICULTURAL PRODUCTIVITY IN SOUTH AFRICA

Cecilia Bellora
Élodie Blanc
Jean-Marc Bourgeon
Eric Strobl

Working Paper 23496
http://www.nber.org/papers/w23496

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2017

We would like to thank Eyal Frank and the participants of the NBER conference on 'Understanding Productivity Growth in Agriculture' for their valuable comments. The research leading to these results has received funding from the European Union's Seventh Framework Programme under Grant Agreement n° 290693 (FOODSECURE) and by the MIT Joint Program on the Science and Policy of Global Change (for a complete list of sponsors and U.S. government funding sources of this program, see http://globalchange.mit.edu/sponsors/). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Estimating the Impact of Crop Diversity on Agricultural Productivity in South Africa
Cecilia Bellora, Élodie Blanc, Jean-Marc Bourgeon, and Eric Strobl
NBER Working Paper No. 23496
June 2017
JEL No. Q15,Q57

**ABSTRACT**

Crop biodiversity has the potential to enhance resistance to strains due to biotic and abiotic factors and to improve crop production and farm revenues. To investigate the effect of crop biodiversity on crop productivity, we build a probabilistic model based on ecological mechanisms to describe crop survival and productivity according to diversity. From this analytic model, we derive reduced forms that are empirically estimated using detailed field data of South African agriculture combined with satellite derived data. Our results confirm that diversity has a positive and significant impact on crop survival odds. We show the consistency of these results with the underlying ecologic and agricultural mechanisms.

Cecilia Bellora
CEPII
75007 Paris, France
cecilia.bellora@cepii.fr

Élodie Blanc
Massachusetts Institute of Technology
77 Massachusetts Ave
Cambridge MA 02139
eblanc@mit.edu

Jean-Marc Bourgeon
INRA and Ecole Polytechnique
Agroparistech
75005 Paris, France.
bourgeon@agroparistech.fr

Eric Strobl
AMSE-GREQAM
Aix-Marseille Université
13205 Marseille Cedex 1
France
eric.strobl@polytechnique.edu

# 1 Introduction

Diversity plays a key role in the resilience to external stresses of farm plants and animals. In particular, crop species diversity increases productivity and production stability (Tilman et al., 2005; Tilman & Downing, 1994; Tilman et al., 1996) in the sense that the probability to find at least one individual that resists to an adverse meteorological phenomenon (for example a drought or a heatwave), or pests and diseases, increases with the diversity within a population. Furthermore, the larger is an homogeneous population, the larger is the number of parasites that use this population as a host and therefore the larger is the probability of a lethal infection (Pianka, 1999). Diversity also allows for species complementarities and, as a consequence, a more efficient use of natural resources (Loreau & Hector, 2001). In short, crop biodiversity has the potential to enhance resistance to strains due to biotic and abiotic factors and to improve crop production and farm revenues.

For these reasons, after the large development of monocultures in the last decades, crop biodiversity is making a comeback. During the last century, farming activities specialized on the most productive crops, in particular in developed countries and in large areas of emerging economies. The decrease in crop biodiversity resulted in increased pest attacks (Landis et al., 2008) and has been compensated by the heavy use of agrochemicals. Nevertheless, chemicals generate negative externalities, irreversible in many cases, on water and soil quality, on wildlife and on human health (Pimentel, 2005; Foley et al., 2011; Jiguet et al., 2012; Beketov et al., 2013), which engender large economic costs (Gallai et al., 2009; Sutton et al., 2011). One of the main challenges for the future is to drastically reduce externalities while satisfying an increasing and changing food demand (Gouel & Guimbard, 2017). In this context, crop biodiversity is more and more seen as a promising way to raise, or at least maintain, agricultural yields while decreasing the use of chemicals (McDaniel et al., 2014). However, more estimations of the actual impacts of crop biodiversity on agricultural yields are needed to build solutions for farmers and to adopt relevant public policies. To this end, we empirically investigate the role of crop biodiversity on crop productivity. We build a probabilistic model based on ecological mechanisms to describe crop survival and productivities according to diversity. From this analytic model, we derive reduced forms that are estimated using data on South African agriculture.

Our results contribute to the existing literature in three main ways. First, we confirm that diversity has a positive and significant impact on produced quantities. An increase in biodiversity is equivalent to a third of the benefits of a comparable increase in irrigation, where irrigation is known to be an important impediment to crop productivity in South Africa, due to unreliable precipitation. Previous empirical investigations on the role of biodiversity on production has produced sometimes contrasted results. Positive impact of biodiversity is found by Di Falco & Chavas (2006) and Carew et al. (2009) in wheat production, in Italy and Canada respectively. Smale et al. (1998) also focus on wheat yield and find a positive impact of biodiversity in rain fed regions of Pakistan, while in irrigated areas higher concentration on few varieties is associated with higher yields. Second, we adopt an approach based on ecology literature, while previous contributions used pure econometric methods, mainly moment based approaches, Di Falco & Chavas (2006, 2009) stressing the crucial role played by skewness in addition to mean and variance. In these cases, the functional forms are disconnected from the ecology literature and therefore do not allow to go into deep details on the way biodiversity impacts productivity. In the economic literature, models of endogenous interaction between biodiversity and crop production have been developed in theoretical papers that analyze the role and the value of biodiversity against specialization on the most productive crops (Weitzman, 2000; Brock & Xepapadeas, 2003; Bellora & Bourgeon, 2016) but have never been coupled with empirical investigations. In contrast, we build a probabilistic model that makes explicit the relationship between biodiversity, biotic and abiotic factors that affect agricultural production. It represents how biodiversity impacts agricultural production. Stochastic shocks affecting agricultural production are endogenous, in accordance with ecology findings.This model can easily be linked to data and grounds our analysis on findings of ecology studies. This approach can also be extended to account for non crop biodiversity (pastures, fallow land, non cultivated areas…), which appears to also play a key role (Tscharntke et al., 2005), and to characterize the impacts on production variability. Third, we draw from the increasingly available satellite data (Donaldson & Storeygard, 2016) to build a rich dataset allowing us to estimate the impact of biodiversity on crop productivity based on our probabilistic model. Normalized Difference Vegetation Index (NDVI) derived from the SPOT 5 satellite images, coupled with land use classification, allow to quantify the crop biomass produced on nearly 65,000 fields covering around 6.5 million hectares in South Africa. We quantify biodiversity using an index taken from the ecological literature, based on species richness (i.e. the total number of species) and their relative abundance, the Shannon index (Shannon, 1948). This index captures the fact that biodiversity is high when the total number of species is large and the distribution of their relative abundances is homogeneous. We are then able to quantify the impacts of inter-specific diversity on the productivity of various crops, while previous studies mainly looked at genetic diversity (i.e. intra-specific diversity of a single crop). We confirm that biodiversity has mainly a local impact: biodiversity is a significant predictor of crop productivity on perimeters having a radius smaller than 2km.

In the remainder of the paper, the theoretical model that motivates our empirical investigations is developed in section 2 and section 3 details its empirical implementation. Then, the database on South African agriculture is presented in section 4. In section 5, we empirically investigate the impact of crop biodiversity on crop production.

# 2 The model

A very robust stylized fact in ecology describes the impact of biotic factors on agricultural production: the more area is dedicated to the same crop, the more pests specialize on this crop and the higher is the frequency of their attacks (Pianka, 1999). Relying on this stylized fact, we build a general probabilistic model of crop production where crops are affected by both abiotic (i.e. weather, water availability, soil properties...) and biotic (i.e. pests) factors causing pre-harvest losses.[1] More precisely, we consider that the total agricultural production depends on the survival probability of each crop, which is directly linked to the probability of a pest attack. The frequency at which pest attacks occur is linked to the way crops are produced: the more diverse are crops, the lower is the probability of a pest attack, the higher is the survival probability and therefore the higher is the expected agricultural production. To describe the diffusion of pests, or equivalently the survival probabilities, we follow the literature in ecology and plant physiology and adopt a beta-binomial distribution, which is usual to depict spatial distributions that are not random but clustered, patchy or heterogeneous (Hughes & Madden, 1993; Shiyomi et al., 2000; Chen et al., 2008; Bastin et al., 2012; Irvine & Rodhouse, 2010).

We assume that a region (or a country) produces $Z$ different crops on $I$ fields of the same size, each field being sowed with one crop only. Characteristics of field $i$ are gathered in vector $X_i = (x_{i1}, \ldots, x_{iK})$ and are related both to abiotic factors and biotic factors. In particular, $X_i$ contains information on the way crops are cultivated (irrigation but also soil quality and field location) and on biodiversity conditions. Depending on the crop cultivated, each field is divided in $n(z)$ patches that are subject to potential lethal strains due, for example, to adverse meteorological conditions or pathogens. We suppose that a patch on field $i$ is destroyed with probability $1 - \lambda_i$ from one (or several) adverse condition, and that otherwise it produces the potential yield $a(z)$ independently of the fate of the other patches on field $i$ or elsewhere.[2] With $n(z)$ patches, the probability of

[1]Losses due to biotic factors can be significant. Oerke (2006) finds that, during the 2001–2003 period, without crop protection, losses in major crops due to pests were comprised between 50% and 80%, at the world level. Thanks to crop protection, they fall between 29% and 37%. Similar results are found for the US by Fernandez-Cornejo et al. (1998).

[2]Obviously, this is a strong assumption. Pests and/or weather do not necessarily totally destroy a patch, but rather affect the quantity of biomass produced. But, in order to maintain tractability, we

having $t$ patches within field $i$ unaffected (and thus $n(z)-t$ destroyed) follows a binomial distribution,

$$\Pr\{\widetilde{T}_i = t | z, \lambda_i\} = \binom{n(z)}{t} \lambda_i^t (1-\lambda_i)^{n(z)-t}$$

where $\widetilde{T}_i$ is the random variable that corresponds to the number of patches that are indeed harvested among the $n(z)$ patches of field $i$ sowed with crop $z$. We consider that the survival probability of the patches of a field, $\lambda_i$, is identically and independently distributed across patches. However, this probability may vary across fields of the same crop (we generally have $\lambda_i \neq \lambda_j$ for any couple of fields $(i,j)$ sowed with the same crop): it depends on natural conditions but also on the characteristics $X_i$ of the field. More precisely, the survival probability of patches on a given field is a draw from a Beta distribution given by

$$\Pr(\lambda_i = \lambda | X_i, z) = \frac{\Gamma[S_{ui}(z) + S_{di}(z)]}{\Gamma[S_{ui}(z)]\,\Gamma[S_{di}(z)]} \lambda^{S_{ui}(z)-1}(1-\lambda)^{S_{di}(z)-1}$$

where $\Gamma(\cdot)$ is the the gamma function, $S_{ui}(z) \equiv e^{\gamma(z)+\boldsymbol{\theta}_u(z).X_i}$ and $S_{di}(z) \equiv e^{\beta(z)+\boldsymbol{\theta}_d(z).X_i}$, $\gamma(z)$ and $\beta(z)$ being positive parameters that determine the randomness of the survival probability of a patch of crop $z$ absent any field specific effect, and the vectors $\boldsymbol{\theta}_u(z) = \{\theta_{uk}(z)\}_{k=1,\ldots,K}$ and $\boldsymbol{\theta}_d(z) = \{\theta_{dk}(z)\}_{k=1,\ldots,K}$ capturing the influence of each field specific effect $X_i$ on the survival probability of crop $z$. The expected number of patches among $n(z)$ that are harvested on field $i$ is given by $\mathrm{E}[\widetilde{T}_i | X_i, z] = n(z)\psi(z, X_i)$ where

$$\psi(z, X_i) = \mathrm{E}[\lambda_i | X_i, z] = \frac{S_{ui}(z)}{S_{ui}(z) + S_{di}(z)} \tag{1}$$

is the expected probability that a particular patch of field $i$ of crop $z$ is harvested given its characteristics $X_i$. Absent field specific effects ($\boldsymbol{\theta}_u(z) = \boldsymbol{\theta}_d(z) = 0$), the expected resilience of a particular stand of crop is given by $\exp\gamma(z)/(\exp\gamma(z) + \exp\beta(z))$. An increase in coefficient $\theta_{uk}(z)$ increases this resilience, while an increase in $\theta_{dk}(z)$ diminishes it, the extent of these effects depending on the corresponding field characteristics $x_{ik}$. The variance of the number of harvested patches on field $i$ is given by $\sigma_i^2 = n(z)V(z, X_i)$ where

$$V(z, X_i) = \psi(z, X_i)[1 - \psi(z, X_i)]\{1 + [n(z) - 1]\rho(z, X_i)\} \tag{2}$$

with

$$\rho(z, X_i) = [1 + S_{ui}(z) + S_{di}(z)]^{-1}.$$

Equation (2) corresponds to the variance of the survival probability of one patch on a field with characteristics $X_i$. Compared to the Bernoulli distribution, (2) contains an

consider that a patch is either unaffected either totally destroyed, rather than partially affected, by adverse conditions. Thus, our random variable is the number of harvested patches rather than the share of biomass that is lost on each patch.

additional term that accounts for the correlation between patches induced by the common distribution of the survival probability, the correlation coefficient being given by $\rho(z, X_i)$.

The production on field $i$ is given by $\widetilde{Y}_i = a(z)\widetilde{T}_i$. It can be equivalently written as

$$\widetilde{Y}_i = \mathrm{E}[\widetilde{Y}_i](1 + \tilde{\varepsilon}_i) \tag{3}$$

where $\mathrm{E}[\widetilde{Y}_i] = a(z)n(z)\psi(z, X_i)$ and $\tilde{\varepsilon}_i = (\widetilde{T}_i - \mathrm{E}[\widetilde{T}_i])/\mathrm{E}[\widetilde{T}_i]$ has a mean equal to 0 and a variance given by

$$\sigma^2_{\varepsilon_i} = \frac{1 - \psi(z, X_i)}{\psi(z, X_i)} \left( \frac{1}{n(z)} + \frac{n(z) - 1}{n(z)} \rho(z, X_i) \right) \approx \frac{1 - \psi(z, X_i)}{\psi(z, X_i)} \rho(z, X_i) \tag{4}$$

when $n(z)$ is large. This variance is mainly due to the correlation between patches on a field that share the same survival probability, captured by $\rho(z, X_i)$. Indeed, $\lambda_i$ follows a beta distribution but the parameters of the distribution depend on the field characteristics $\mathbf{X_i}$ and are thus different across fields. In other words, with a sufficiently large number of patches on each field, the difference in the quantities produced is mainly driven by field characteristics.

This simple ecological model of crop production can thus be summarized as follows: the number of patches that are harvested on field $i$, $\widetilde{T}_i$, follows a beta-binomial distribution determined by the parameters $\gamma(z), \beta(z), \theta_{uk}, \theta_{dk}$. Parameters $\theta_{uk}$ and $\theta_{dk}$ determine the impact of the $k^{th}$ field characteristic $x_{ik}$ on $\widetilde{T}_i$, in addition to the parameters $\gamma(z)$ and $\beta(z)$ that are shared by all fields that grow crop $z$. Depending on the values of $\theta_{uk}$ and $\theta_{dk}$, each characteristic $x_{ik}$ can increase or decrease the expected number of harvested patches on field $i$ and skew the distribution of $\widetilde{T}_i$ to the right or to the left, modifying the probability of extreme events like the loss of all the patches in a field.

In the following section, we build an empirical strategy to estimate the impact of the characteristics of a field on the distribution of $\widetilde{T}_i$. In particular, we are interested in the impact on crop production of the crop biodiversity surrounding the field considered and expect this impact to be positive, according to findings and mechanisms described in the ecology literature.

# 3 Empirical strategy

Starting from the probabilistic model, our aim is to estimate the parameters $\theta_u(z)$, $\theta_d(z)$, $\gamma(z)$ and $\beta(z)$ of the distribution of the survival probability $\widetilde{T}$. We first have to derive from each field production the corresponding survival probability $\hat{\lambda}_i$. They are obtained by dividing the production level by the potential maximum production $a(z)n(z)$ level. This potential production is not observed in practice, it is derived in the following from the maximum observed production level $Y_M(z) \equiv \max Y_i(z)$ using $a(z)n(z) = (1 + \alpha)Y_M(z)$

where $\alpha \geq 0$.[3] With a linear regression of the equation

$$\ln\left(\frac{\hat{\lambda}_i}{1-\hat{\lambda}_i}\right) = \delta + \boldsymbol{\Delta}.X_i \tag{5}$$

for each type of crop, we obtain the estimate $\hat{\delta}(z)$ of $\gamma(z) - \beta(z)$ and $\hat{\boldsymbol{\Delta}}(z)$ of $\boldsymbol{\theta}_u(z) - \boldsymbol{\theta}_d(z)$. This first regression estimates the contribution of biodiversity (and other field characteristics) to the ratio of survival and death probabilities. Coefficients $\Delta$ show the variation of the growth rate of the odds associated with a marginal increase in each explanatory variable. These first results are interesting *per se* but also allow to derive an expected patch survival rate for each field $i$ using

$$\hat{\psi}_i = 1/(1 + e^{-\hat{\delta}(z) - \hat{\boldsymbol{\Delta}}(z).X_i})$$

and a serie of dispersion values

$$\hat{\varepsilon}_i = (\hat{\lambda}_i - \hat{\psi}_i)/\hat{\psi}_i.$$

From (4) which can be written as

$$\sigma^2_{\varepsilon_i} = \frac{1 - \psi(z, X_i)}{\psi(z, X_i) + S_{ui}}$$

we get, solving for $S_{ui}$

$$S_{ui} = \frac{1 - \psi(z, X_i)(1 + \sigma^2_{\varepsilon_i})}{\sigma^2_{\varepsilon_i}}.$$

As $S_{ui} = \exp(\gamma + \boldsymbol{\theta_u}(z).X_i)$, we construct the variable

$$\hat{Z}_i = \frac{1 - \hat{\psi}_i(1 + \hat{\varepsilon}_i^2)}{\hat{\varepsilon}_i^2}$$

and we perform an OLS estimation of the equation

$$\ln(\hat{Z}_i) = \gamma(z) + \boldsymbol{\theta}_u(z).X_i \tag{6}$$

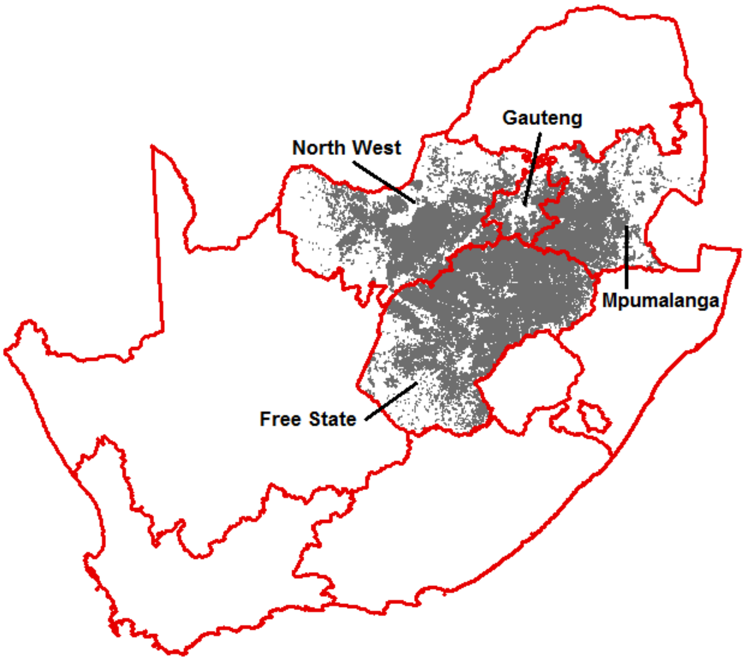
to obtain $\hat{\gamma}(z)$ and $\hat{\boldsymbol{\theta}}_u(z)$. We then get $\hat{\beta}(z) = \hat{\gamma}(z) - \hat{\delta}(z)$ and $\hat{\boldsymbol{\theta}}_d(z) = \hat{\boldsymbol{\theta}}_u(z) - \hat{\boldsymbol{\Delta}}(z)$.

# 4 Data

Combining different data sources, we construct a very detailed original database on South African agriculture that quantifies the production and describes the characteristics of a

[3] In the following, we consider $\alpha = 0.5$. Robustness checks for $\alpha = 0.1$ are available in the Appendix.

Figure 1: Localization of the considered fields in South Africa



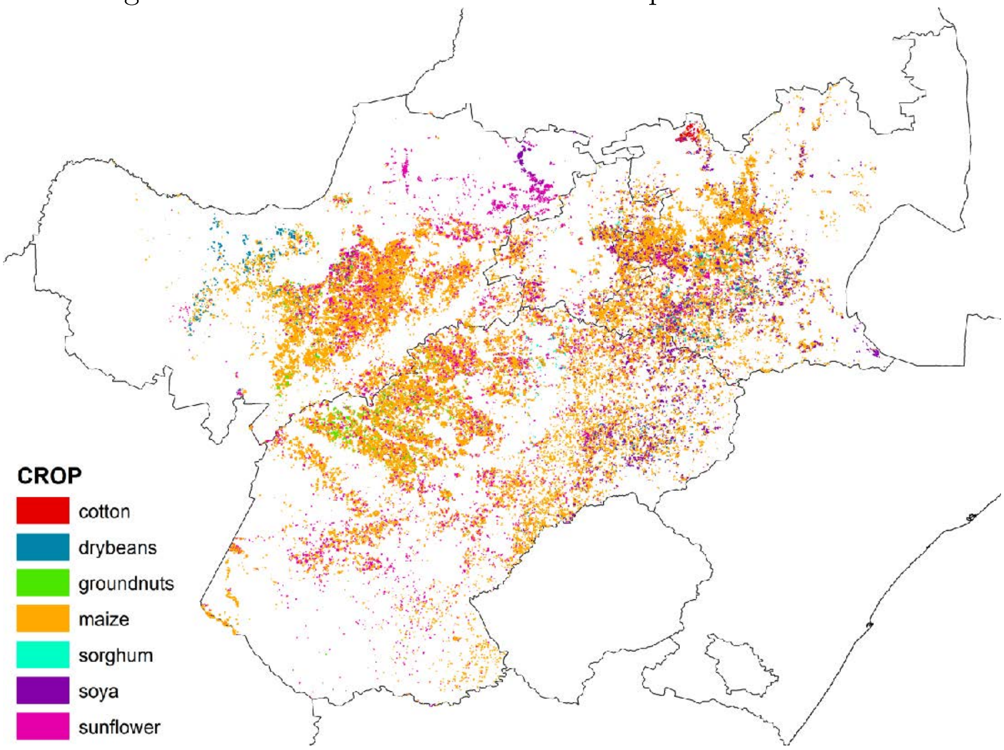

very large number of fields using satellite data. First, field boundaries are identified, then agricultural production is characterised on each field by identifying the crops that are grown and measuring the biomass produced. Field characteristics are then collected, concerning in particular water balance, length of the growing season and crop interspecific biodiversity.

## 4.1 Crop fields

Field boundaries, available for South African provinces of Free State, Gauteng, North West and Mpumalanga, are determined using the Producer Independent Crop Estimate System (PICES) which combines satellite imagery, Geographic Information System (GIS), point frame statistical platforms, and aerial observations (Ferreira et al., 2006). Satellite imagery of cultivated fields is obtained from the SPOT 5 satellite at a 2.5-m resolution. Plot boundaries are then digitised using GIS and field cloud covered polygons are removed before processing. Over the four regions of interest, PICES distinguishes circa 280 000 fields covering an area of around 6.5 million hectares. To approximately match the resolution of the crop production indicator we use (see 4.2), which is only available at the 250m resolution, the analysis is limited to fields larger than 6.25ha. Additionally, we exclude pasture and fallow land. This restricts the sample to 64,682 fields. Figure 1 presents the location of the considered crop fields in South Africa. While the summary statistics in Table 1 show that fields are on average about 28.4 hectares, the large standard deviation (24.8) indicates that they vary substantially in size (the largest field is 720 hectares large).

Using the digitized satellite images described above, the Agricultural Geo-referenced

Figure 2: Distribution of the studied crops in South Africa




Information System (AGIS) developed by the South African Department of Agriculture provides information on the crop cultivated on each field. To do so, sample points were selected randomly and surveyed by trained observers from a very light aircraft in order to determine crop type (Ferreira et al., 2006). Crop information collected during the aerial surveys on the sample points was subsequently used as a training set for crop type classification for each field and for accuracy assessment. These estimated crop classifications were then checked against a producer based survey for the Gauteng region. The Gauteng census survey showed that less than 1.8% of crop types had been misclassified. All in all, seven summer crops were distinguished for the provinces of Free State, Gauteng, North West and Mpumalanga for the summer season 2006/2007: cotton, dry beans, groundnuts, maize, sorghum, soybean and sunflower. An example of the distribution of crop types is provided in figure 2. The summary statistics for the entire sample in Table 2 show that maize was the dominant crop cultivated in the three provinces: maize fields represent nearly 70 per cent of the total number of fields we consider. Other important crops were sunflower and soybean, standing at 15 and 11 per cent each. In contrast, all other crop types constituted less than 2 per cent individually. One should note that even if one were to adjust the crop type shares by their areas, a similar ranking remains, with a slight redistribution of shares towards the smaller crop types. For instance, the share of maize dropped to 62 per cent of the total crop area.

The AGIS crop boundaries dataset also provides information regarding irrigation, from

Table 1: Plot summary statistics

| Variable | Mean | Standard deviation | Variable | Mean | Standard deviation |
|---|---|---|---|---|---|
| All crops | | | Maize | | |
| NDVI | 0.61 | 0.13 | NDVI | 0.61 | 0.13 |
| Water balance | −41.07 | 11.73 | Water balance | −42.60 | 11.89 |
| Season length (days) | 129.64 | 35.48 | Season length (days) | 126.98 | 35.33 |
| Plot area(ha) | 28.36 | 23.47 | Plot area(ha) | 22.72 | 2.04 |
| Farm area (ha) | 315.2 | 500.00 | Farm area (ha) | 193.63 | 3.53 |
| Irrigation (%) | 5.05 | – | Irrigation (%) | 4.94 | 21.67 |
| Cotton | | | Sorghum | | |
| NDVI | 0.73 | 0.05 | NDVI | 0.69 | 0.10 |
| Water balance | −28.61 | 14.54 | Water balance | −33.54 | 7.70 |
| Season length (days) | 130.94 | 52.00 | Season length (days) | 125.20 | 26.17 |
| Plot area(ha) | 15.63 | 1.66 | Plot area(ha) | 19.27 | 1.91 |
| Farm area (ha) | 184.41 | 2.32 | Farm area (ha) | 45.63 | 2.79 |
| Irrigation (%) | 56.60 | 49.80 | Irrigation (%) | 1.96 | 13.87 |
| Dry bean | | | Soybean | | |
| NDVI | 0.62 | 0.14 | NDVI | 0.71 | 0.09 |
| Water balance | −38.96 | 12.39 | Water balance | −32.18 | 7.59 |
| Season length (days) | 145.58 | 38.66 | Season length (days) | 129.19 | 30.14 |
| Plot area(ha) | 21.43 | 2.03 | Plot area(ha) | 19.09 | 1.91 |
| Farm area (ha) | 68.50 | 3.10 | Farm area (ha) | 95.98 | 2.95 |
| Irrigation (%) | 3.75 | 19.00 | Irrigation (%) | 5.89 | 23.55 |
| Groundnuts | | | Sunflower | | |
| NDVI | 0.52 | 0.11 | NDVI | 0.59 | 0.13 |
| Water balance | −55.56 | 6.34 | Water balance | −44.20 | 12.10 |
| Season length (days) | 115.40 | 34.67 | Season length (days) | 131.27 | 38.01 |
| Plot area(ha) | 28.61 | 2.03 | Plot area(ha) | 20.54 | 2.01 |
| Farm area (ha) | 73.91 | 2.96 | Farm area (ha) | 80.59 | 3.14 |
| Irrigation (%) | 4.10 | 19.84 | Irrigation (%) | 7.80 | 26.81 |

Table 2: Distribution of the considered crops

| Crop | Nb. of fields | Share of total fields (%) | Share of total area (%) |
|---|---|---|---|
| Dry beans | 1,227 | 1.88 | 1.88 |
| Groundnuts | 1,292 | 2.5 | 2.50 |
| Maize | 45,256 | 69.77 | 72.27 |
| Sorghum | 715 | 1.10 | 0.93 |
| Soybean | 6,825 | 10.52 | 8.80 |
| Sunflower | 9,441 | 14.56 | 13.51 |
| Cotton | 106 | 0.16 | 0.10 |
| Total | 64,862 | 100.00 | 100.00 |

which only 5% of the fields considered benefit (Table 1).

Finally, all fields can be linked to their respective farms with a unique farm identifier. In total the fields were owned by 12,462 different farms, where on average each farm was proprietor of 5 fields. However, ownership differed substantially, with the largest ownership gathering 193 fields, and 3,704 single field farms.

## 4.2 Crop production measure

We estimate crop biomass production using the satellite derived Normalised Difference Vegetation Index (NDVI). Vegetation indices provide consistent spatial and temporal representations of vegetation conditions, when locally derived information is not available. As a matter of fact, numerous studies have demonstrated that NDVI values are significantly correlated with biomass production, and therefore yields, of various crops, including wheat (Das et al., 1993; Gupta et al., 1993; Doraiswamy & Cook, 1995; Hochheim & Barber, 1998; Labus et al., 2002), sorghum (Potdar, 1993), maize (Hayes & Decker, 1996; Prasad et al., 2006), rice (Nuarsa et al., 2011; Quarmby et al., 1993), soybean (Prasad et al., 2006), barley (Weissteiner & Kuhbauch, 2005), millet (Groten, 1993) and tomato (Koller & Upadhaya, 2005). Moreover, NDVI has also been shown to provide a very good indicator of crop phenological development (Benedetti & Rossini, 1993).

The NDVI index is calculated using ratios of vegetation spectral reflectance over incoming radiation in each spectral band. The NDVI data are extracted from the MOD13Q1 dataset,[4] which gathers reflectance information collected by the MODerate-resolution Imaging Spectroradiometer (MODIS) instrument operating on NASA's Terra satellite (Huete et al., 2002). From these data, NDVI can be formulated as:

$$NDVI = \frac{NIR - VIS}{NIR + VIS}$$

where the difference between near-infrared reflectance (NIR) and visible reflectance (VIS) values is normalised by the total reflectance and varies between - 1 and 1 (Eidenshink, 1992). The more biomass is produced, the more the NDVI is close to 1. Negative and very low values corresponding to water and barren areas were excluded from the analysis by design. Nevertheless, NDVI has some limitations. In particular, it enters an asymptotic regime for high values of biomass. It reaches its maximum when leaves totally cover the soil and does not allow to distinguish between dense or very dense vegetation, contrary to other vegetation indices that do not saturate over densely vegetated regions (Huete et al., 1997). In that sense, NDVI is less reliable to estimate the biomass production of dense vegetation, like forest. However, it is very sensible to photosynthetic activity and therefore remains highly indicative of the biomass produced in cultivated fields. Carl-

[4] Available online from https://lpdaac.usgs.gov/lpdaac/content/view/full/6652

son & Ripley (1997) precisely describe the asymptotic regime of NDVI and Ma et al. (2001) confirm this analysis and relate biomass produced to NDVI using the following relationship, extrapolated for soybean:

$$Y = d + bNDVI^c \tag{7}$$

where $Y$ represents the quantities produced (or the yield), $d$, $b$ and $c$ are three parameters. The only parameter needed in the following is $c$, taken equal to 4.54, following Ma et al. (2001).[5] Denoting by

$$N_i = NDVI_i - NDVI_0 \tag{8}$$

with $NDVI_0 = |d/b|^{1/c}$, $bN_i^c$ gives an estimation of the quantities produced on field $i$, $Y(i)$.[6]

Crop growing seasons are characterized by the planting date and the phenology cycle, which determines the length of the season. In South Africa, planting generally occurs between October and December in order to reduce the vulnerability to erratic precipitation (Ferreira et al., 2006). However, phenology cycles, and hence growing seasons, can differ substantially among crop types and even for fields of the same crop type. In order to take account of this, we used the TIMESAT program[7] (Jönsson & Eklundh, 2002, 2004) to determine crop and field specific growing seasons. We are then able to approximate the start and end of growing seasons based on distribution properties of the NDVI. Summary statistics in Table 1 show that growing seasons are on average 130 days, with a standard deviation of 35 days.

Finally, as is standard in the literature of satellite derived plant growth measures, we use the maximum NDVI over the growing season as an indicator of crop production (Zhang et al., 2006). It takes on an average value of 0.61 with a standard deviation of 0.13 (see Table 1).

## 4.3 Crop water balance

An important determinant of crop growth is water availability. A common simple proxy for it is the difference between rainfall and the evaporative demand of the air, i.e, evapotranspiration. To calculate this, we use gridded daily precipitation and reference evapotranspiration data taken from the USGS Early Warning Famine climatic database.[8] More specifically, daily rainfall data, given at the 0.1 degree resolution (approximately 11 km),

[5]We take the estimate coming from the regression showing the best fit on data used by Ma et al. (2001).

[6]For value smaller than $NDVI_0$, the produced quantities are equal to 0, the NDVI capturing the light reflected by the bare soil.

[7]The algorithm within the TIMESAT software is commonly used to extract seasonality information from satellite time-series data.

[8]http://earlywarning.usgs.gov/fews

are generated with the rainfall estimation algorithm RFE (version 2.0) dataset implemented by the National Oceanic and Atmospheric Administration (NOAA) - Climate Prediction Center (CPC) using a combination of rain gauges and satellite observations. Daily reference evapotranspiration data, available at a 1 degree resolution (approximately 111km), were calculated using a 6-hourly assimilation of conventional and satellite observational data of air temperature, atmospheric pressure, wind speed, relative humidity and solar radiation extracted from the National Oceanic and Atmospheric Administration Global Data Assimilation System. Using these gridded data each field was then assigned a daily precipitation and potential evapotranspiration value over its growing season to then calculate out its average daily water balance. The mean and standard deviation of this measure are given in Table 1.

## 4.4 Biodiversity index

Among field characteristics, we are particularly interested in crop biodiversity. Diversity measures, extensively used in biology and ecology literature, take into account specie richness (i.e. the number of species present) and evenness (i.e. the distribution of species). In the following, we quantify biodiversity at the field level adopting one of the most widely used indicators, the Shannon index (Shannon, 1948):

$$H_\ell = -\sum_z B_\ell(z) \ln B_\ell(z) \tag{9}$$

where the subscript $i$ refers to a specific field, $\ell$ defines the size of the perimeter considered as relevant, and $B_\ell(z)$ is the proportion of area within perimeter $\ell$ that is of crop $z$ type. $H_{i\ell}$ is then calculated for a given perimeter $\ell$, defined by its radius, applied to the centroid of the field $i$ considered. The more diverse crops are and the more equal their abundances, the larger is the Shannon index. When all crops are equally common, all $B(z)$ values will equal $1/Z$ ($Z$ being the total number of crops) and $H$ will be equal to $\ln Z$. On the contrary, the more unequal the abundances of the crops are, the smaller is the index, approaching 0 (and being equal to 0 if $Z = 1$). With respect to other common indicators, like the Simpson's index,[9] the Shannon index is known to put less weight on the more abundant species and to be more sensitive to differences in total species richness and in changes in populations showing small relative abundances (Baumgärtner, 2006). In our specification, the distance threshold for the radius $\ell$ is 0.75km; the distance is then increased 250m by 250m to reach 3km, the maximum distance considered. We provide summary statistics for the Shannon index in Table 3. Widening the perimeter under consideration increases the value of the Shannon index substantially. For example, the 3 km index is nearly 5 times larger than the 0.75km index. This strongly suggests that

[9] With our notations, the Simpson's index is given by $1 - \sum_z B_\ell^2(z)$.

Table 3: Summary statistics for the Shannon index

| | All crops | | Dry bean | | Groundnuts | | Maize | | Sorghum | | Soybean | | Sunflower | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\ell$ | $\overline{H}$ | $\sigma_H$ | $\overline{H}$ | $\sigma_H$ | $\overline{H}$ | $\sigma_H$ | $\overline{H}$ | $\sigma_H$ | $\overline{H}$ | $\sigma_H$ | $\overline{H}$ | $\sigma_H$ | $\overline{H}$ | $\sigma_H$ |
| 0.75km | 0.03 | 0.14 | 0.17 | 0.31 | 0.21 | 0.31 | 0.08 | 0.21 | 0.11 | 0.25 | 0.17 | 0.29 | 0.14 | 0.26 |
| 1.00km | 0.06 | 0.18 | 0.27 | 0.37 | 0.35 | 0.35 | 0.14 | 0.26 | 0.20 | 0.31 | 0.28 | 0.33 | 0.23 | 0.31 |
| 1.25km | 0.07 | 0.21 | 0.36 | 0.39 | 0.43 | 0.35 | 0.19 | 0.29 | 0.28 | 0.35 | 0.37 | 0.35 | 0.30 | 0.33 |
| 1.50km | 0.09 | 0.23 | 0.45 | 0.41 | 0.48 | 0.34 | 0.23 | 0.30 | 0.35 | 0.36 | 0.44 | 0.36 | 0.35 | 0.34 |
| 1.75km | 0.10 | 0.24 | 0.50 | 0.42 | 0.51 | 0.34 | 0.27 | 0.31 | 0.40 | 0.37 | 0.49 | 0.35 | 0.40 | 0.34 |
| 2.00km | 0.12 | 0.25 | 0.55 | 0.43 | 0.53 | 0.32 | 0.31 | 0.32 | 0.45 | 0.37 | 0.54 | 0.35 | 0.43 | 0.33 |
| 2.25km | 0.13 | 0.26 | 0.60 | 0.43 | 0.55 | 0.31 | 0.33 | 0.32 | 0.49 | 0.37 | 0.58 | 0.34 | 0.46 | 0.33 |
| 2.50km | 0.13 | 0.27 | 0.63 | 0.43 | 0.56 | 0.30 | 0.36 | 0.32 | 0.52 | 0.37 | 0.61 | 0.33 | 0.48 | 0.32 |
| 2.75km | 0.14 | 0.28 | 0.66 | 0.43 | 0.57 | 0.30 | 0.38 | 0.31 | 0.56 | 0.36 | 0.63 | 0.32 | 0.50 | 0.32 |
| 3.00km | 0.15 | 0.28 | 0.68 | 0.43 | 0.58 | 0.29 | 0.40 | 0.31 | 0.59 | 0.36 | 0.65 | 0.32 | 0.52 | 0.31 |

*Note:* The table reports the mean ( $\overline{H}$) and the standard deviation ($\sigma_H$) of the distribution of the Shannon index, measured for the different crops considered, on different perimeters, characterized by their radius, $\ell$.

crop types are strongly spatially agglomerated, and thus locally less diverse.

# 5 Empirical analysis

Our first empirical task is to investigate whether biodiversity affects crop field production. To this end, we rely on the strategy defined in section 3. In short, we build data on crop production using (7) and (8). We use them to calculate the survival probability in each field, $\hat{\lambda}_i$. Then, with a linear regression on specification (5), we estimate the impact of biodiversity on the odds, i.e. the ratio of the probability for a given field to survive to the probability of death.

Crop productivity depends not only on crop biodiversity but also on more general natural conditions (weather, season length…), field attributes (irrigation, area…) and farm management attributes (pesticides, mechanization, economies of scale…). Therefore, the vector of control variables $\mathbf{X}$ includes crop fixed effects, crop water balance ($WB$) and its squared value ($WB^2$), an irrigation dummy indicator ($IR$), the season length ($SEAS\_LENGTH$), the logarithm of the field area in hectares ($\ln(AREA)$), the latitude ($LAT$) and longitude ($LON$) of the centroid of the field, the percentage of cropland within a defined perimeter that is irrigated ($PC\_AREA\_IR$), and the percentage of land devoted to the same crop that belongs to the same farm, within a defined perimeter ($PC\_AREA\_FARM$). We also include farm fixed effects to capture crop management techniques that are common within farms as well as farm wide economies of scale. Crop specific dummies allow us to control for the fact that different crops will have different vegetation growth intensity as captured by satellite reflectance data.

The results of the regression on equation (5) for all crops pooled are presented in Table 4. In the first column, we simply include our field specific control variables (vector $X$). The first column shows results for a perimeter defined by a radius $\ell$ equal to 0.75km. As can be seen, crop water balance has a significant positive and exponentially increasing impact on the survival rate of crops. However, having an irrigation system acts more to increase the survival rate of crops and therefore fields' productivity. It also makes crops less reliant on water balance (in a linear fashion) as would be expected. The coefficient on season length suggests that the longer the season lasts, the lower the crop survival rate is. In other words, the longer the season, the higher the probabilities that an adverse event affects crops. Larger fields have lower survival rates than smaller ones. Finally, being located more in the East results in crop survival probability, possibly because of more favorable climatic or soil conditions, while being further South or North is inconsequential for field productivity within our sample.

If we consider now the degree of crop diversity, as measured by the Shannon index, we observe that an increase in surrounding biodiversity improves the survival ratio in a given field, and consequently its productivity. Arguably, however, our diversity index may just be capturing the fact that neighboring areas are different in ways that are correlated with the diversity of crops. To take account of these factors, we thus control for the percentage of the surrounding area that is irrigated and the percentage of the surrounding area of fields of the same crop type that belongs to the same farm.

When increasing the defined perimeter to calculate the Shannon index to 1 km, adjusting the variables $PC\_AREA\_IR$ and $PC\_AREA\_FARM$ in an analogous fashion, the impact of crop biodiversity on survival rate remains statistically significant, but decreases by 26%. As far as control variables are concerned, the share of area irrigated unequivocally increases the biomass production while the share of area belonging to the same farm within the perimeter we consider seems to have no significant impact on the biomass production. Further increasing the perimeter similarly continues to produce a significant positive impact of biodiversity, the coefficient increasing by 40% per cent. However, when further expanding the threshold of our definition of the relevant neighborhood, biodiversity still acts as a significant predictor of survival probability but its contribution decreases and finally disappears for a perimeter's radius greater than 2km.[10] This suggests that biodiversity is relatively locally defined, i.e. within less than 2km, but likely close to 1.25km.

We then look at the heterogeneity of impacts across crops, considering sequentially each of the six crops for which data are available (cotton is not considered in the regressions by crop since the available sample – 106 fields, 0.16% of the total available fields and 0.1 of the total cropland considered – is too small). The results show that, on the one hand, biodiversity has a significant impact on survival probability of maize, soybean and sunflower, and that the relevant perimeter size of the biodiversity index depends on the crop. Table 5 also reveals that biodiversity has no significant impact on dry bean,

[10] We also experimented with increasing the perimeter up to 10km, but the coefficient on $H$ remains insignificant in all cases.

Table 4: Regression results, all crops pooled

| Variables | $\ell = 0.75$km | $\ell = 1.00$km | $\ell = 1.25$km | $\ell = 1.50$km | $\ell = 1.75$km | $\ell = 2.00$km | $\ell = 2.25$km | $\ell = 2.50$km | $\ell = 2.75$km | $\ell = 3.00$km |
|---|---|---|---|---|---|---|---|---|---|---|
| H | 0.03597*** | 0.026725** | 0.037328*** | 0.018801** | 0.016833** | 0.013968* | 0.001571 | 0.002857 | −0.004031 | −0.010203 |
| | (0.011643) | (0.010443) | (0.010458) | (0.00719) | (0.008218) | (0.007961) | (0.009092) | (0.00829) | (0.009999) | (0.009274) |
| WB | 0.005932*** | 0.005862*** | 0.005809*** | 0.005779*** | 0.005852*** | 0.005946*** | 0.006065*** | 0.006057*** | 0.006077*** | 0.006133*** |
| | (0.00135) | (0.001368) | (0.001354) | (0.001348) | (0.001356) | (0.001363) | (0.00136) | (0.001368) | (0.001357) | (0.001365) |
| $WB^2$ | 0.000085*** | 0.000084*** | 0.000084*** | 0.000083*** | 0.000084*** | 0.000085*** | 0.000086*** | 0.000086*** | 0.000086*** | 0.000087*** |
| | (0.000016) | (0.000016) | (0.000016) | (0.000016) | (0.000016) | (0.000016) | (0.000016) | (0.000016) | (0.000016) | (0.000016) |
| WB×IR | −0.009296*** | −0.009562*** | −0.009641*** | −0.009712*** | −0.009637*** | −0.009533*** | −0.009538*** | −0.009543*** | −0.009598*** | −0.009598*** |
| | (0.002855) | (0.002858) | (0.002811) | (0.002768) | (0.002787) | (0.002809) | (0.002784) | (0.002789) | (0.002814) | (0.002834) |
| $WB^2$×IR | −0.000012 | −0.000015 | −0.000014 | −0.000016 | −0.000017 | −0.000016 | −0.000017 | −0.000017 | −0.000017 | −0.000018 |
| | (0.000039) | (0.000039) | (0.000039) | (0.000038) | (0.000038) | (0.000039) | (0.000038) | (0.000038) | (0.000038) | (0.000039) |
| SEAS_LEN | −0.002714*** | −0.002725*** | −0.002727*** | −0.002742*** | −0.002738*** | −0.002738*** | −0.002739*** | −0.002737*** | −0.002736*** | −0.002734*** |
| | (0.000183) | (0.000183) | (0.000185) | (0.000188) | (0.000188) | (0.000188) | (0.000187) | (0.000187) | (0.000186) | (0.000185) |
| ln(AREA) | −0.028502*** | −0.029762*** | −0.030119*** | −0.030802*** | −0.030816*** | −0.031277*** | −0.031213*** | −0.031429*** | −0.03156*** | −0.031693*** |
| | (0.004043) | (0.004023) | (0.003988) | (0.003915) | (0.003949) | (0.003928) | (0.003984) | (0.003959) | (0.003971) | (0.004008) |
| IR | 0.316261*** | 0.309933*** | 0.305125*** | 0.309627*** | 0.31731*** | 0.323508*** | 0.327262*** | 0.328253*** | 0.328347*** | 0.329908*** |
| | (0.050474) | (0.050908) | (0.051096) | (0.051218) | (0.051214) | (0.051173) | (0.051046) | (0.050637) | (0.050879) | (0.05092) |
| LON | 0.729432*** | 0.732371*** | 0.743884*** | 0.740554*** | 0.738677*** | 0.730207*** | 0.723389*** | 0.725916*** | 0.725441*** | 0.722365*** |
| | (0.145315) | (0.144919) | (0.14452) | (0.145287) | (0.145463) | (0.145262) | (0.145221) | (0.144333) | (0.144138) | (0.144694) |
| LAT | −0.22436 | −0.228158 | −0.230148 | −0.226988 | −0.225607 | −0.226386 | −0.227417 | −0.222639 | −0.218449 | −0.2201 |
| | (0.262772) | (0.263929) | (0.265465) | (0.277887) | (0.280737) | (0.28579) | (0.286228) | (0.293318) | (0.296958) | (0.296455) |
| PC_AREA_IR | 0.122233*** | 0.14442*** | 0.179287*** | 0.191377*** | 0.182695*** | 0.177287*** | 0.148563*** | 0.147581*** | 0.147332*** | 0.119077*** |
| | (0.016362) | (0.017688) | (0.01541) | (0.018342) | (0.021349) | (0.020165) | (0.023612) | (0.021937) | (0.026856) | (0.026089) |
| PC_AREA_FARM | 0.01385** | 0.009282 | 0.004797 | −0.004242 | −0.001923 | −0.00738 | −0.004304 | −0.013599 | −0.018975 | −0.021637 |
| | (0.006303) | (0.006278) | (0.007849) | (0.009018) | (0.008449) | (0.009257) | (0.011217) | (0.013014) | (0.014318) | (0.016453) |

*Note:* ***, ** and * indicate 1, 5 and 10 per cent significance levels, respectively. Robust standard errors are in parentheses, clustered at field level. Farm and crop fixed effects are included but not reported. Sample: 64,862 fields, 12,462 farms. Results for individual crops are reported in the Appendix.

Table 5: Impact of biodiversity on the odds of survival probabilities

| $\ell$ | Pooled crops | Dry Beans | Ground Nuts | Maize | Soya | Sunflower |
|---|---|---|---|---|---|---|
| 0.75 | 0.036*** | −0.0168 | 0.0584 | 0.0493*** | −0.0067 | 0.0291 |
| | (0.0116) | (0.0685) | (0.0902) | (0.0169) | (0.021) | (0.0332) |
| 1.00 | 0.0267** | −0.0023 | −0.0454 | 0.0301** | 0.0052 | 0.0459* |
| | (0.0104) | (0.0591) | (0.093) | (0.0117) | (0.0215) | (0.0272) |
| 1.25 | 0.0373*** | 0.0404 | −0.0538 | 0.0315*** | 0.0477** | 0.0257 |
| | (0.0105) | (0.0557) | (0.1051) | (0.0116) | (0.0222) | (0.0241) |
| 1.50 | 0.0188** | 0.0768 | −0.2042* | 0.008 | 0.0574** | −0.0105 |
| | (0.0072) | (0.0673) | (0.1102) | (0.0094) | (0.0222) | (0.0314) |
| 1.75 | 0.0168** | 0.0498 | −0.091 | 0.0111 | 0.0848*** | −0.0091 |
| | (0.0082) | (0.0717) | (0.116) | (0.0101) | (0.0217) | (0.033) |
| 2.00 | 0.014* | −0.0428 | 0.0496 | 0.0084 | 0.096*** | −0.0308 |
| | (0.008) | (0.0714) | (0.1142) | (0.0091) | (0.026) | (0.0362) |
| 2.25 | 0.0016 | 0.0006 | −0.0159 | −0.0004 | 0.0683** | −0.062** |
| | (0.0091) | (0.0695) | (0.1347) | (0.0105) | (0.0259) | (0.0311) |
| 2.50 | 0.0029 | 0.0312 | −0.0244 | 0.0071 | 0.0647*** | −0.0912*** |
| | (0.0083) | (0.0735) | (0.1765) | (0.0084) | (0.023) | (0.0313) |
| 2.75 | −0.004 | 0.0245 | −0.0196 | 0.0069 | 0.0573** | −0.119*** |
| | (0.01) | (0.0798) | (0.1836) | (0.0105) | (0.0253) | (0.0444) |
| 3.00 | −0.0102 | −0.0328 | −0.0348 | 0.0041 | 0.0547* | −0.1558*** |
| | (0.0093) | (0.0744) | (0.1864) | (0.0121) | (0.028) | (0.0459) |
| Fields | 64,682 | 1,227 | 1,292 | 45,256 | 6,825 | 9,441 |

*Note:* ***, ** and * indicate 1, 5 and 10 per cent significance levels, respectively. Robust standard errors are in parentheses, clustered at field level. Farm and crop fixed effects are included but not reported. Sample: 64,862 fields, 12,462 farms. The table shows the coefficients for the variable $H_\ell$ for each of the six crops considered. Complete regression results are reported in the annex.

groundnuts and sorghum. This can be explained by the fact that each of the latter crops represents less than 2% of the total number of fields. In other words, the area dedicated to these crops is small, the fields are probably sufficiently scattered to not suffer from the proliferation of their pests. Therefore, the biodiversity variation on the perimeter that we consider has a negligible marginal effect on the biomass production. When looking at the crops which biomass production is affected by crop biodiversity, we see that the relevant perimeter for biodiversity varies: biodiversity has a positive and significant impact on the production of maize only for perimeters equal to or smaller than 1.25km whereas the relevant perimeter for soybean is equal to or greater than between 1.25km. Surprisingly, biodiversity has a negative and significant impact on sunflower biomass production on perimeters with a radius larger than 2.25km; a positive significant impact is found only for $\ell$ equal to 1.00km. This heterogeneity is probably linked to the fact that pests responsible for biomass losses differ among the the three crops we consider. Generally, the main potential crop losses are caused by weeds but, thanks to the improvement in weed control techniques, the main actual losses come from animals (mainly insects) and pathogens (Oerke, 2007). More precisely, in South Africa, maize is mainly attacked by insects (DAFF, 2014a), while sunflower and soybean are mainly attacked by diseases caused by fungi and viruses (DAFF, 2009, 2014b).

The impact of irrigation also varies and depends on crop characteristics, as shown in tables 8, 9 and 10. Maize is one of the most efficient cultivated plants in South Africa as far as water use is concerned (DAFF, 2014a), hence a positive and significant impact of irrigation. On the contrary, sunflower is highly inefficient in water-use and, as well as soybean, is mostly rain-fed grown.[11] This could explain the absence of a significant impact of irrigation on biomass production for these crops. Finally, soybean biomass production is positively affected by the size of the field while sunflower survival rate is inversely related to field size and maize is unaffected. These effects could be related to plant physiology or to higher mechanization allowed by larger fields and having a positive impact on the final yield of soybean.

These results confirm the positive impact of crop biodiversity on agricultural production and underlines its heterogeneity across crops, with sunflower being an exception. Furthermore, it is important to note that we estimate biodiversity effects in the presence of pesticides, for which we do not totally control. Indeed, farm fixed effects capture practices that are common to all the fields within the same farm and crop fixed effects capture practices common to all crops, but the level of pesticides actually applied remains unknown. Then, the effects we observe can be considered as residual. The positive impact of biodiversity on crop survival, only second to the one of irrigation and more generally water management, is all the more important in that respect. Even when pesticides are possibly applied, biodiversity still has the capacity to improve crop survival rate.

The results presented above detail the impact of crop biodiversity on crop survival rates. Our approach through a probabilistic model can be used to add a step to disentangle more precisely the mechanisms at stake. In particular, the parameters of the beta-binomial distribution of the survival probability of fields can be estimated, following the approach detailed in section 3. We perform an OLS regression on equation (6) to directly estimates the value of $\boldsymbol{\theta}_d$; $\boldsymbol{\theta}_u$ is given by the difference between the coefficients found in the linear regressions on equations (5) and (6).

Results are presented in Table 6, which reports the values of parameters $\theta_u$ and $\theta_d$ for all the explanatory variables for selected values of $\ell$, and Table 7 which shows the values of the parameters for the Shannon index, computed on all the possible perimeters. As it is visible from Table 6, in practice significant individual values for the parameters of the beta-binomial distribution can be found on in a limited number of cases. In particular, it is interesting to note that biodiversity has a positive impact on the survival rate of maize by increasing $S_u$ more than $S_d$ (see equation (1)), while the positive impact found for sunflower comes from a larger decrease in $S_d$ than in $S_u$. This difference in mechanisms

[11] Soybean is mostly rain-fed grown because of low profitability and difficult water management. Indeed, water shortage is critical during the pod set stage while excessive water supply prior or after the flowering may jeopardize the final yield.

at stake confirms the important role played by crop specificities (plant physiology as well as predominant pests) on the possible impacts of crop biodiversity on agricultural production.

# 6 Conclusion

Using a new large database built from satellite imagery, we confirm that crop biodiversity has a positive impact on agricultural production, which is heterogeneous across crops, sunflower being an exception. Maintaining a large diversity of crops in the landscape increases agricultural production level. These impacts, that were previously described at regional scales, are robust when we consider a larger area. We show the consistency of these results with the underlying ecologic and agricultural mechanisms. For this purpose, we build a probabilistic model in which stochastic factors linked to biodiversity, namely pests, are endogenous, as it is shown in the ecology literature, while previous results were derived using functional forms arbitrarily chosen.

In the absence of data on pesticides use, their effects are not precisely measured in this model, which only evaluates the residual effects of biodiversity. However, our approach can be easily extended to pesticides. This would have the advantage of measuring their effects not on an isolated field, but rather within a varied set of agricultural productions. Nevertheless, our analysis shows that residual effects are important and that a better spatial distribution of crops could lead to a significant improvement in crop yields. This could be achieved if farmers distribute their crops on their farms to take account of these effects not only on their own yields but also taking into account their surroundings, which supposes that they coordinate.

Describing the mechanisms governing the impact of biodiversity on crop survival, our model can also be extended to consider wild biodiversity. Indeed, maintaining uncultivated small areas in agricultural landscapes is considered to diminish pests attacks. Adding data on uncultivated areas to our dataset, the contribution of these initiatives could be easily evaluated.

Furthermore, enriching the dataset, in particular with data on pesticide use, could help to precisely estimate the parameters of the beta-binomial distribution of survival probabilities. Characterising the distribution could bring elements on the impacts of crop biodiversity on the variance and the skewness of the distribution, i.e. on the probability of extreme events, in particular the complete loss of the harvest. These results are rarely analysed in the literature (Di Falco & Chavas, 2009), while they are particularly relevant for farmers.

Notwithstanding these limitations, our results confirm that crop diversification can be seen as a possible strategy to increase agricultural productivity or to maintain its level

Table 6: Estimated values of the parameters of the distribution of the survival probabilities – All explanatory variables, for selected perimeters

| | Pooled crops ($\ell = 1.25$km) | | Maize ($\ell = 1.25$km) | | Soya ($\ell = 2.00$km) | | Sunflower ($\ell = 1.00$km) | |
|---|---|---|---|---|---|---|---|---|
| Exp. variable | $\theta_u$ | $\theta_d$ | $\theta_u$ | $\theta_d$ | $\theta_u$ | $\theta_d$ | $\theta_u$ | $\theta_d$ |
| H | 0.1409*** | 0.10357*** | 0.14552*** | 0.11399*** | 0.04782 | −0.04816 | −0.23772** | −0.28358*** |
| WB | −0.00419 | −0.00999 | −0.00447 | −0.00679 | −0.03697* | −0.04193** | −0.00652 | −0.01731 |
| $WB^2$ | −0.00008 | −0.00016** | −0.00005 | −0.00011 | −0.00048 | −0.00064 | −0.00002 | −0.00011 |
| WB × IR | 0.00207 | 0.01171 | 0.00234 | 0.00667 | 0.02111 | 0.02403 | −0.03593* | −0.01164 |
| $WB^2$ × IR | −0.00004 | −0.00002 | 0.00006 | 0.00006 | 0.0005 | 0.00044 | −0.0006** | −0.00034 |
| SEAS_LEN | 0.0005 | 0.00322*** | 0.00113* | 0.0042*** | −0.00367*** | −0.00052 | 0.00104 | 0.0022** |
| ln(AREA) | −0.01987 | 0.01024 | −0.07582*** | −0.04862** | 0.01812 | −0.01301 | −0.06639 | 0.02386 |
| IR | 0.10136 | −0.20376 | −0.29752 | −0.81894** | 0.40038 | 0.25032 | −0.05214 | −0.31581 |
| LON | 0.09129 | −0.65259 | 1.12783 | 0.5228 | 2.83914* | 1.16488 | −5.42781*** | −7.41688*** |
| LAT | −0.55793 | −0.32778 | −0.49671 | −0.55656 | −1.82428 | −0.94141 | −1.32601 | 0.94822 |
| PC_AREA_IR | −0.04723 | −0.22652*** | −0.00232 | −0.23696*** | −0.11008 | −0.216 | 0.17849 | 0.22753 |
| PC_AREA_FARM | 0.03263 | 0.02783 | 0.04747* | 0.04235 | 0.20894** | 0.19909** | −0.10167 | −0.11259 |

*Note:* ***, ** and * indicate 1, 5 and 10 per cent significance levels, respectively. Standard t-tests are used for $\theta_d$, estimated with equation (5), and a Z-test is used for $\theta_u$, calculated using the coefficients produced by equations (5) and (6), as detailed in section 3. The values of $\ell$ presented correspond to the largest and most significant impact of crop biodiversity on the survival rate.

Table 7: Estimated values of the parameters of the distribution of the survival probabilities – Biodiversity parameters

| | $\ell$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | 0.75 | 1.00 | 1.25 | 1.50 | 1.75 | 2.00 | 2.25 | 2.50 | 2.75 | 3.00 |
| Maize | | | | | | | | | | |
| $\gamma$ | −0.135** | −0.146** | −0.183*** | −0.165** | −0.181*** | −0.167** | −0.151** | −0.155** | −0.153** | −0.173*** |
| $\beta$ | 0.219*** | 0.207*** | 0.169** | 0.188*** | 0.172** | 0.187*** | 0.203*** | 0.2*** | 0.202*** | 0.182*** |
| $\theta_u$ | 0.083 | 0.156*** | 0.146*** | 0.137*** | 0.155*** | 0.109** | 0.094** | 0.087** | 0.069 | 0.027 |
| $\theta_d$ | 0.033 | 0.125*** | 0.114*** | 0.129*** | 0.144*** | 0.1** | 0.094** | 0.08* | 0.062 | 0.022 |
| Soya | | | | | | | | | | |
| $\gamma$ | 0 | −0.02 | −0.042 | −0.036 | −0.058 | −0.041 | −0.042 | −0.045 | −0.051 | −0.067 |
| $\beta$ | 0.119 | 0.098 | 0.075 | 0.082 | 0.06 | 0.077 | 0.077 | 0.074 | 0.068 | 0.053 |
| $\theta_u$ | 0.303*** | 0.24*** | 0.07 | −0.019 | 0.128 | 0.048 | −0.009 | 0.047 | −0.009 | 0.087 |
| $\theta_d$ | 0.31*** | 0.235*** | 0.022 | −0.076 | 0.043 | −0.048 | −0.077 | −0.018 | −0.067 | 0.033 |
| Sunflower | | | | | | | | | | |
| $\gamma$ | −0.114 | −0.114 | −0.153** | −0.133* | −0.148* | −0.129 | −0.122 | −0.122 | −0.118 | −0.131 |
| $\beta$ | −0.036 | −0.037 | −0.077 | −0.055 | −0.071 | −0.052 | −0.043 | −0.044 | −0.039 | −0.053 |
| $\theta_u$ | −0.217 | −0.238** | −0.142 | −0.123 | −0.15 | −0.067 | −0.197* | −0.284** | −0.272 | −0.355** |
| $\theta_d$ | −0.246* | −0.284*** | −0.168 | −0.112 | −0.141 | −0.036 | −0.135 | −0.193 | −0.153 | −0.199 |

*Note:* ***, ** and * indicate 1, 5 and 10 per cent significance levels, respectively. Standard t-tests are used for $\theta_d$, estimated with equation (5), and a Z-test is used for $\theta_u$, calculated using the coefficients produced by equations (5) and (6), as detailed in section 3. $\theta_u$ and $\theta_d$ presented here are those related to the Shannon index.

while decreasing the use of pesticides.

# References

Bastin, G., Scarth, P., Chewings, V., Sparrow, A., Denham, R., Schmidt, M., O'Reagain, P., Shepherd, R., & Abbott, B. (2012). Separating grazing and rainfall effects at regional scale using remote sensing imagery: a dynamic reference-cover method. *Remote Sensing of Environment*, 121(0), 443–457.

Baumgärtner, S. (2006). *Measuring the diversity of what? And for what purpose? A conceptual comparison of ecological and economic biodiversity indices.* Working paper, Department of Economics, University of Heidelberg.

Beketov, M. A., Kefford, B. J., Schafer, R. B., & Liess, M. (2013). Pesticides reduce regional biodiversity of stream invertebrates. *Proceedings of the National Academy of Sciences*, 110(27), 11039–11043.

Bellora, C. & Bourgeon, J.-M. (2016). *Agricultural trade, biodiversity effects and food price volatility.* Working paper 2016-06, CEPII.

Benedetti, R. & Rossini, P. (1993). On the use of NDVI profiles as a tool for agricultural statistics: the case study of wheat yield estimate and forecast in Emilia Romagna. *Remote Sensing of Environment*, 45(3), 311–326.

Brock, W. A. & Xepapadeas, A. (2003). Valuing biodiversity from an economic perspective: A unified economic, ecological, and genetic approach. *American Economic Review*, 93(5), 1597–1614.

Carew, R., Smith, E. G., & Grant, C. (2009). Factors influencing wheat yield and variability: evidence from Manitoba, Canada. *Journal of Agricultural and Applied Economics*, 41(3), 625–639.

Carlson, T. N. & Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover and leaf area index. *Remote Sensing of Environment*, 62(3), 241–252.

Chen, J., Shiyomi, M., Bonham, C., Yasuda, T., Hori, Y., & Yamamura, Y. (2008). Plant cover estimation based on the beta distribution in grassland vegetation. *Ecological Research*, 23(5), 813–819.

DAFF (2009). *Sunflower.* Production guideline, Department of Agriculture, Forestry and Fisheries, Republic of South Africa.

DAFF (2014a). *Maize.* Production guideline, Department of Agriculture, Forestry and Fisheries, Republic of South Africa.

DAFF (2014b). *Soybeans.* Production guideline, Department of Agriculture, Forestry and Fisheries, Republic of South Africa.

Das, D. K., Mishra, K. K., & Kalra, N. (1993). Assessing growth and yield of wheat using remotely-sensed canopy temperature and spectral indices. *International Journal of Remote Sensing*, 14(17), 3081–3092.

Di Falco, S. & Chavas, J. (2006). Crop genetic diversity, farm productivity and the management of environmental risk in rainfed agriculture. *European Review of Agricultural Economics*, 33(3), 289–314.

Di Falco, S. & Chavas, J. (2009). On crop biodiversity, risk exposure, and food security in the highlands of Ethiopia. *American Journal of Agricultural Economics*, 91(3), 599–611.

Donaldson, D. & Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, 30(4), 171–198.

Doraiswamy, P. C. & Cook, P. W. (1995). Spring wheat yield assessment using NOAA AVHRR data. *Canadian Journal of Remote Sensing*, 21, 43–51.

Eidenshink, J. C. (1992). The 1990 conterminous U.S. AVHRR dataset. *Photogrammetric Engineering and Remote Sensing*, 58(6), 809–813.

Fernandez-Cornejo, J., Jans, S., & Smith, M. (1998). Issues in the economics of pesticide use in agriculture: a review of the empirical evidence. *Review of Agricultural Economics*, 20(2), 462–488.

Ferreira, S. L., Newby, T., & du Preez, E. (2006). Use of remote sensing in support of crop area estimates in South Africa. In *ISPRS WG VIII/10 Workshop 2006, Remote Sensing Support to Crop Yield Forecast and Area Estimates*.

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockstrm, J., Sheehan, J., Siebert, S., Tilman, D., & Zaks, D. P. M. (2011). Solutions for a cultivated planet. *Nature*, 478(7369), 337–342.

Gallai, N., Salles, J.-M., Settele, J., & Vaissière, B. E. (2009). Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecological Economics*, 68(3), 810–821.

Gouel, C. & Guimbard, H. (2017). *Nutrition transition and the structure of global food demand*. Working Papers 2017-05, CEPII.

Groten, S. M. E. (1993). NDVI – Crop monitoring and early yield assessment of Burkina Faso. *International Journal of Remote Sensing*, 14(8), 1495–1515.

Gupta, R., Prasad, S., Rao, G., & Nadham, T. (1993). District level wheat yield estimation using NOAA/AVHRR NDVI temporal profile. *Advances in Space Research*, 13(5), 253–256.

Hayes, M. J. & Decker, W. L. (1996). Using NOAA AVHRR data to estimate maize production in the United States Corn Belt. *International Journal of Remote Sensing*, 17(16), 3189–3200.

Hochheim, K. & Barber, D. G. (1998). Spring wheat yield estimation for Western Canada using NOAA NDVI data. *Canadian journal of remote sensing*, 24, 17–27.

Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., & Ferreira, L. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1-2), 195–213.

Huete, A., Liu, H. Q., Batchily, K., & Van Leeuwen, W. (1997). A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment*, 59(3), 440–451.

Hughes, G. & Madden, L. (1993). Using the beta-binomial distribution to describe aggregated patterns of disease incidence. *Phytopathology*, 83(7), 759–763.

Irvine, K. M. & Rodhouse, T. J. (2010). Power analysis for trend in ordinal cover classes: implications for long-term vegetation monitoring. *Journal of Vegetation Science*, 21(6), 1152–1161.

Jiguet, F., Devictor, V., Julliard, R., & Couvet, D. (2012). French citizens monitoring ordinary birds provide tools for conservation and ecological sciences. *Acta Oecologica*, 44(0), 58–66.

Jönsson, P. & Eklundh, L. (2002). Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing*, 40(8), 1824–1832.

Jönsson, P. & Eklundh, L. (2004). TIMESAT – A program for analyzing time-series of satellite sensor data. *Computers & Geosciences*, 30(8), 833–845.

Koller, M. & Upadhaya, S. K. (2005). Prediction of processing tomato yield using a crop growth model and remotely sensed aerial images. *Transactions of the ASAE*, 48(6), 2335–2341.

Labus, M. P., Nielsen, G. A., Lawrence, R. L., Engel, R., & Long, D. S. (2002). Wheat yield estimates using multi-temporal NDVI satellite imagery. *International Journal of Remote Sensing*, 23(20), 4169–4180.

Landis, D. A., Gardiner, M. M., van der Werf, W., & Swinton, S. M. (2008). Increasing corn for biofuel production reduces biocontrol services in agricultural landscapes. *Proceedings of the National Academy of Sciences*, 105(51), 20552–20557.

Loreau, M. & Hector, A. (2001). Partitioning selection and complementarity in biodiversity experiments. *Nature*, 412(6842), 72–76.

Ma, B. L., Dwyer, L. M., Costa, C., Cober, E. R., & Morrison, M. J. (2001). Early prediction of soybean yield from canopy reflectance measurements. *Agronomy Journal*, 93(6), 1227–1234.

McDaniel, M. D., Tiemann, L. K., & Grandy, A. S. (2014). Does agricultural crop diversity enhance soil microbial biomass and organic matter dynamics? a meta-analysis. *Ecological Applications*, 24(3), 560–570.

Nuarsa, I. W., Nishio, F., Nishio, F., Hongo, C., & Hongo, C. (2011). Relationship between rice spectral and rice yield using MODIS data. *Journal of Agricultural Science*, 3(2).

Oerke, E. (2006). Crop losses to pests. *Journal of Agricultural Science*, 144(1), 31–43.

Oerke, E. (2007). *Encyclopedia of Pest Management, Volume 2*, chapter Crop losses to animal pests, plant pathogens and weeds, (pp. 116 – 120). CRC Press.

Pianka, E. R. (1999). *Evolutionary Ecology (6th edition)*. Benjamin Cummings.

Pimentel, D. (2005). Environmental and economic costs of the application of pesticides primarily in the United States. *Environment, Development and Sustainability.*

Potdar, M. B. (1993). Sorghum yield modelling based on crop growth parameters determined from visible and near-IR channel NOAA AVHRR data. *International Journal of Remote Sensing*, 14(5), 895–905.

Prasad, A. K., Chai, L., Singh, R. P., & Kafatos, M. (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation*, 8(1), 26–33.

Quarmby, N. A., Milnes, M., Hindle, T. L., & Silleos, N. (1993). The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction. *International Journal of Remote Sensing*, 14(2), 199–210.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423, 623–656.

Shiyomi, M., Takahashi, S., & Yoshimura, J. (2000). A measure for spatial heterogeneity of a grassland vegetation based on the beta-binomial distribution. *Journal of Vegetation Science*, 11(5), 627–632.

Smale, M., Hartell, J., Heisey, P. W., & Senauer, B. (1998). The contribution of genetic resources and diversity to wheat production in the Punjab of Pakistan. *American Journal of Agricultural Economics*, 80(3), 482–493.

Sutton, M. A., Oenema, O., Erisman, J. W., Leip, A., van Grinsven, H., & Winiwarter, W. (2011). Too much of a good thing. *Nature*, 472(7342), 159–161.

Tilman, D. & Downing, J. A. (1994). Biodiversity and stability in grasslands. *Nature*, 367(6461), 363–365.

Tilman, D., Polasky, S., & Lehman, C. (2005). Diversity, productivity and temporal stability in the economies of humans and nature. *Journal of Environmental Economics and Management*, 49(3), 405–426.

Tilman, D., Wedin, D., & Knops, J. (1996). Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature*, 379(6567), 718–720.

Tscharntke, T., Klein, A. M., Kruess, A., Steffan-Dewenter, I., & Thies, C. (2005). Landscape perspectives on agricultural intensification and biodiversity – ecosystem service management. *Ecology Letters*, 8(8), 857–874.

Weissteiner, C. J. & Kuhbauch, W. (2005). Regional yield forecasts of malting barley (Hordeum vulgare L.) by NOAA-AVHRR remote sensing data and ancillary data. *Journal of Agronomy and Crop Science*, 191(4), 308–320.

Weitzman, M. L. (2000). Economic profitability versus ecological entropy. *The Quarterly Journal of Economics*, 115(1), 237–263.

Zhang, X., Friedl, M. A., & Schaaf, C. B. (2006). Global vegetation phenology from moderate resolution imaging spectroradiometer (MODIS): evaluation of global patterns and comparison with in situ measurements. *Journal of Geophysical Research*, 111(G4).

# Appendices

Table 8: Regression results, maize

| Variables | $\ell = 0.75$km | $\ell = 1.00$km | $\ell = 1.25$km | $\ell = 1.50$km | $\ell = 1.75$km | $\ell = 2.00$km | $\ell = 2.25$km | $\ell = 2.50$km | $\ell = 2.75$km | $\ell = 3.00$km |
|---|---|---|---|---|---|---|---|---|---|---|
| H | 0.049324*** | 0.03006** | 0.031536*** | 0.008002 | 0.011067 | 0.008432 | −0.000411 | 0.007078 | 0.006927 | 0.004122 |
| | (0.01694) | (0.01172) | (0.01162) | (0.00942) | (0.01013) | (0.00907) | (0.01052) | (0.00838) | (0.01052) | (0.01206) |
| WB | 0.00256 | 0.002349 | 0.002324 | 0.002345 | 0.002588 | 0.002664 | 0.002719 | 0.002673 | 0.002716 | 0.002792 |
| | (0.00205) | (0.00211) | (0.00213) | (0.00212) | (0.00212) | (0.00212) | (0.00211) | (0.00212) | (0.00212) | (0.00211) |
| $WB^2$ | 0.000058** | 0.000055** | 0.000055* | 0.000055* | 0.000057** | 0.000058** | 0.000058** | 0.000058** | 0.000058** | 0.000059** |
| | (0.00003) | (0.00003) | (0.00003) | (0.00003) | (0.00003) | (0.00003) | (0.00003) | (0.00003) | (0.00003) | (0.00003) |
| WB × IR | −0.004 | −0.004005 | −0.004331 | −0.004487 | −0.004498 | −0.004365 | −0.004338 | −0.004272 | −0.004348 | −0.004355 |
| | (0.00381) | (0.00381) | (0.00385) | (0.00382) | (0.00381) | (0.0038) | (0.00381) | (0.00383) | (0.00382) | (0.00382) |
| $WB^2$ × IR | 0.00001 | 0.000009 | 0.000006 | 0.000003 | 0.000002 | 0.000003 | 0.000004 | 0.000005 | 0.000004 | 0.000004 |
| | (0.00005) | (0.00005) | (0.00005) | (0.00005) | (0.00005) | (0.00005) | (0.00005) | (0.00005) | (0.00005) | (0.00005) |
| SEAS_LEN | −0.003065*** | −0.003076*** | −0.003074*** | −0.00309*** | −0.00309*** | −0.003091*** | −0.003091*** | −0.003086*** | −0.003086*** | −0.003084*** |
| | (0.00022) | (0.00022) | (0.00022) | (0.00022) | (0.00022) | (0.00022) | (0.00022) | (0.00022) | (0.00022) | (0.00022) |
| ln(AREA) | −0.025124*** | −0.026647*** | −0.027206*** | −0.027764*** | −0.027999*** | −0.028617*** | −0.028494*** | −0.028824*** | −0.029087*** | −0.029314*** |
| | (0.00502) | (0.00505) | (0.00503) | (0.005) | (0.00497) | (0.00494) | (0.00496) | (0.005) | (0.00502) | (0.00506) |
| IR | 0.525926*** | 0.526102*** | 0.521412*** | 0.5236*** | 0.532595*** | 0.538173*** | 0.540895*** | 0.542285*** | 0.542369*** | 0.542742*** |
| | (0.06782) | (0.06822) | (0.0689) | (0.06808) | (0.06765) | (0.0674) | (0.06772) | (0.06803) | (0.06791) | (0.06785) |
| LON | 0.592476*** | 0.598901*** | 0.605021*** | 0.605076*** | 0.598296*** | 0.590941*** | 0.586532*** | 0.59056*** | 0.589496*** | 0.58852*** |
| | (0.1643) | (0.16441) | (0.16562) | (0.16693) | (0.16726) | (0.1672) | (0.16647) | (0.16604) | (0.16547) | (0.16522) |
| LAT | 0.060083 | 0.056434 | 0.059845 | 0.0614 | 0.063162 | 0.063196 | 0.063251 | 0.067501 | 0.063533 | 0.065777 |
| | (0.29092) | (0.2911) | (0.29468) | (0.30447) | (0.31056) | (0.31875) | (0.31991) | (0.32565) | (0.32867) | (0.33086) |
| PC_AREA_IR | 0.161766*** | 0.200108*** | 0.23464*** | 0.249899*** | 0.189434*** | 0.181045*** | 0.184278*** | 0.17636*** | 0.143832*** | 0.129077*** |
| | (0.02093) | (0.02464) | (0.02847) | (0.03523) | (0.02914) | (0.03225) | (0.02981) | (0.03259) | (0.03697) | (0.04124) |
| PC_AREA_FARM | 0.016634*** | 0.011013 | 0.005122 | −0.002783 | −0.002691 | −0.014197 | −0.013329 | −0.024775* | −0.0321** | −0.041696*** |
| | (0.00632) | (0.00741) | (0.00905) | (0.00872) | (0.00827) | (0.0096) | (0.01173) | (0.01377) | (0.0133) | (0.01439) |

*Note:* ***, ** and * indicate 1, 5 and 10 per cent significance levels, respectively. Robust standard errors are in parentheses, clustered at field level. Farm and crop fixed effects are included but not reported. Sample : 45,256 fields.

Table 9: Regression results, soybean

| Variables | $\ell = 0.75$km | $\ell = 1.00$km | $\ell = 1.25$km | $\ell = 1.50$km | $\ell = 1.75$km | $\ell = 2.00$km | $\ell = 2.25$km | $\ell = 2.50$km | $\ell = 2.75$km | $\ell = 3.00$km |
|---|---|---|---|---|---|---|---|---|---|---|
| H | −0.006713 | 0.005198 | 0.047717** | 0.057405** | 0.084795*** | 0.095981*** | 0.068344** | 0.064653*** | 0.057342** | 0.054712* |
| | (0.02099) | (0.02145) | (0.02223) | (0.02216) | (0.0217) | (0.02604) | (0.02592) | (0.02298) | (0.02528) | (0.02795) |
| WB | 0.004647 | 0.004555 | 0.004655 | 0.004991 | 0.004941 | 0.004951 | 0.004943 | 0.004883 | 0.004856 | 0.004761 |
| | (0.00452) | (0.00445) | (0.00451) | (0.00457) | (0.00454) | (0.00458) | (0.00456) | (0.00454) | (0.00449) | (0.00453) |
| $WB^2$ | 0.000153* | 0.000153* | 0.000154* | 0.000161* | 0.00016* | 0.000158* | 0.000158* | 0.000157* | 0.000156* | 0.000155* |
| | (0.00008) | (0.00008) | (0.00008) | (0.00008) | (0.00008) | (0.00008) | (0.00008) | (0.00008) | (0.00008) | (0.00008) |
| WB × IR | −0.002915 | −0.003133 | −0.003224 | −0.002806 | −0.003384 | −0.002916 | −0.003191 | −0.003484 | −0.003621 | −0.003305 |
| | (0.00673) | (0.00676) | (0.00667) | (0.00678) | (0.00675) | (0.00675) | (0.00675) | (0.00679) | (0.00672) | (0.00672) |
| $WB^2$ × IR | 0.000062 | 0.000058 | 0.000059 | 0.000064 | 0.000053 | 0.000061 | 0.000057 | 4.0.00008 | 0.000046 | 0.000054 |
| | (0.00014) | (0.00014) | (0.00014) | (0.00014) | (0.00014) | (0.00014) | (0.00014) | (0.00014) | (0.00014) | (0.00014) |
| SEAS_LEN | −0.00314*** | −0.003162*** | −0.003155*** | −0.003152*** | −0.00316*** | −0.003154*** | −0.003158*** | −0.003166*** | −0.003162*** | −0.003174*** |
| | (0.0003) | (0.00031) | (0.00031) | (0.00031) | (0.00031) | (0.00031) | (0.00031) | (0.00031) | (0.00031) | (0.00031) |
| ln(AREA) | 0.03411*** | 0.032274*** | 0.032211*** | 0.031126*** | 0.031929*** | 0.031137*** | 0.030721*** | 0.0308*** | 0.03058*** | 0.030458*** |
| | (0.01063) | (0.01043) | (0.01036) | (0.0102) | (0.0101) | (0.01013) | (0.01019) | (0.01031) | (0.01036) | (0.01044) |
| IR | 0.150017 | 0.146435 | 0.138695 | 0.153893 | 0.141014 | 0.150062 | 0.149172 | 0.147687 | 0.145054 | 0.149707 |
| | (0.10415) | (0.10302) | (0.10382) | (0.10569) | (0.10644) | (0.10628) | (0.1063) | (0.106) | (0.10499) | (0.10563) |
| LON | 1.627478*** | 1.638173*** | 1.656017*** | 1.649476*** | 1.640765*** | 1.674261*** | 1.657722*** | 1.653005*** | 1.644894*** | 1.650042*** |
| | (0.47296) | (0.46983) | (0.46969) | (0.4685) | (0.46495) | (0.46398) | (0.47553) | (0.47301) | (0.47177) | (0.47339) |
| LAT | −0.816238 | −0.8566 | −0.871631 | −0.872731 | −0.890484 | −0.882871 | −0.872779 | −0.852793 | −0.819648 | −0.826059 |
| | (0.5594) | (0.56418) | (0.55804) | (0.55554) | (0.54973) | (0.56051) | (0.57276) | (0.57777) | (0.59718) | (0.5912) |
| PC_AREA_IR | 0.039263 | 0.04134 | 0.064699 | 0.002541 | 0.118914* | 0.105912 | 0.048999 | 0.130932* | 0.164139** | 0.105992 |
| | (0.05188) | (0.04648) | (0.05239) | (0.06266) | (0.06465) | (0.0684) | (0.07239) | (0.07116) | (0.08216) | (0.08969) |
| PC_AREA_FARM | 0.046059*** | 0.042829*** | 0.028548* | 0.023905 | 0.030914 | 0.00985 | 0.004308 | 0.012491 | −0.009431 | −0.012533 |
| | (0.01677) | (0.01371) | (0.01668) | (0.02268) | (0.023) | (0.02804) | (0.03143) | (0.03075) | (0.02946) | (0.03543) |

*Note:* ***, ** and * indicate 1, 5 and 10 per cent significance levels, respectively. Robust standard errors are in parentheses, clustered at field level. Farm and crop fixed effects are included but not reported. Sample : 6,825 fields.

Table 10: Regression results, sunflower

| Variables | $\ell = 0.75$km | $\ell = 1.00$km | $\ell = 1.25$km | $\ell = 1.50$km | $\ell = 1.75$km | $\ell = 2.00$km | $\ell = 2.25$km | $\ell = 2.50$km | $\ell = 2.75$km | $\ell = 3.00$km |
|---|---|---|---|---|---|---|---|---|---|---|
| H | 0.029129 | 0.045858* | 0.025659 | −0.010517 | −0.009139 | −0.030796 | −0.062028** | −0.091234*** | −0.11896*** | −0.155816*** |
| | (0.03315) | (0.02716) | (0.02408) | (0.03137) | (0.03304) | (0.03625) | (0.03109) | (0.03132) | (0.04435) | (0.04587) |
| WB | 0.010933*** | 0.010786*** | 0.010705*** | 0.010712*** | 0.010609*** | 0.010616*** | 0.010755*** | 0.01083*** | 0.010893*** | 0.01089*** |
| | (0.00313) | (0.00312) | (0.00312) | (0.00314) | (0.00315) | (0.00315) | (0.00316) | (0.00314) | (0.00314) | (0.00313) |
| $WB^2$ | 0.0001** | 0.000098** | 0.000096** | 0.000096** | 0.000095** | 0.000095** | 0.000097** | 0.000098** | 0.000099** | 0.000099** |
| | (0.00004) | (0.00004) | (0.00004) | (0.00004) | (0.00004) | (0.00004) | (0.00004) | (0.00004) | (0.00004) | (0.00004) |
| WB × IR | −0.024084** | −0.024295** | −0.024257** | −0.024444** | −0.024299** | −0.024283** | −0.024433** | −0.024823** | −0.024869** | −0.024968** |
| | (0.01062) | (0.01082) | (0.01087) | (0.01097) | (0.01099) | (0.01113) | (0.01109) | (0.01088) | (0.01084) | (0.0107) |
| $WB^2$ × IR | −0.000255* | −0.000259* | −0.000252* | −0.000257* | −0.000252* | −0.000252* | −0.000256* | −0.000261* | −0.00026* | −0.000265* |
| | (0.00013) | (0.00013) | (0.00013) | (0.00013) | (0.00014) | (0.00014) | (0.00014) | (0.00014) | (0.00014) | (0.00013) |
| SEAS_LEN | −0.00115** | −0.001162** | −0.001194** | −0.001199** | −0.001207** | −0.001218** | −0.001213** | −0.001209** | −0.001202** | −0.00121** |
| | (0.00047) | (0.00047) | (0.00047) | (0.00047) | (0.00047) | (0.00047) | (0.00047) | (0.00047) | (0.00047) | (0.00047) |
| ln(AREA) | −0.088306*** | −0.090249*** | −0.08991*** | −0.090747*** | −0.090004*** | −0.089171*** | −0.088977*** | −0.089546*** | −0.088721*** | −0.089974*** |
| | (0.01235) | (0.01281) | (0.01241) | (0.01248) | (0.01257) | (0.01247) | (0.01259) | (0.01248) | (0.01254) | (0.01251) |
| IR | 0.277555 | 0.263678 | 0.240256 | 0.245185 | 0.240317 | 0.240486 | 0.243622 | 0.237489 | 0.233546 | 0.240787 |
| | (0.21241) | (0.21366) | (0.2157) | (0.2192) | (0.218) | (0.21886) | (0.21586) | (0.21066) | (0.21079) | (0.20794) |
| LON | 1.996051*** | 1.989068*** | 1.987142*** | 1.953806*** | 1.949349*** | 1.911225*** | 1.883959*** | 1.876782*** | 1.861706*** | 1.871549*** |
| | (0.45091) | (0.45045) | (0.44734) | (0.45238) | (0.44769) | (0.45383) | (0.45432) | (0.45147) | (0.44027) | (0.43685) |
| LAT | −2.277531*** | −2.274225*** | −2.274789*** | −2.252246*** | −2.275891*** | −2.294913*** | −2.246656*** | −2.211773*** | −2.188102*** | −2.137399*** |
| | (0.69584) | (0.69817) | (0.69947) | (0.7078) | (0.70513) | (0.69387) | (0.69355) | (0.69257) | (0.69118) | (0.6974) |
| PC_AREA_IR | −0.13485** | −0.049032 | 0.064285 | 0.045335 | 0.103357 | 0.140709 | 0.101707 | 0.124156 | 0.202023** | 0.093958 |
| | (0.05729) | (0.05677) | (0.0744) | (0.07616) | (0.09062) | (0.09005) | (0.09664) | (0.09717) | (0.10157) | (0.1017) |
| PC_AREA_FARM | 0.037325* | 0.010925 | 0.007846 | 0.004324 | 0.017379 | 0.043618 | 0.053532 | 0.050195 | 0.063014 | 0.070944 |
| | (0.02122) | (0.01888) | (0.028) | (0.03464) | (0.03508) | (0.04194) | (0.04522) | (0.04996) | (0.05248) | (0.05545) |

*Note:* ***, ** and * indicate 1, 5 and 10 per cent significance levels, respectively. Robust standard errors are in parentheses, clustered at field level. Farm and crop fixed effects are included but not reported. Sample : 9,441 fields.

Table 11: Robustness check: regression results, all crops pooled, $\alpha = 0.1$

| Variables | $\ell = 0.75$km | $\ell = 1.00$km | $\ell = 1.25$km | $\ell = 1.50$km | $\ell = 1.75$km | $\ell = 2.00$km | $\ell = 2.25$km | $\ell = 2.50$km | $\ell = 2.75$km | $\ell = 3.00$km |
|---|---|---|---|---|---|---|---|---|---|---|
| H | 0.040329*** | 0.030251*** | 0.041919*** | 0.022412*** | 0.020188** | 0.016991* | 0.003242 | 0.004229 | −0.003499 | −0.010054 |
| | (0.01241) | (0.01132) | (0.01098) | (0.00767) | (0.00885) | (0.00875) | (0.00984) | (0.00894) | (0.01098) | (0.01026) |
| WB | 0.006523*** | 0.006442*** | 0.006385*** | 0.006348*** | 0.006423*** | 0.006529*** | 0.006667*** | 0.006657*** | 0.006678*** | 0.006742*** |
| | (0.00151) | (0.00153) | (0.00152) | (0.00151) | (0.00152) | (0.00153) | (0.00152) | (0.00153) | (0.00152) | (0.00153) |
| $WB^2$ | 0.000091*** | 0.00009*** | 0.00009*** | 0.000089*** | 0.00009*** | 0.000091*** | 0.000092*** | 0.000092*** | 0.000092*** | 0.000093*** |
| | (0.00002) | (0.00002) | (0.00002) | (0.00002) | (0.00002) | (0.00002) | (0.00002) | (0.00002) | (0.00002) | (0.00002) |
| WB × IR | −0.010002*** | −0.010297*** | −0.010385*** | −0.010464*** | −0.010382*** | −0.010264*** | −0.010272*** | −0.010278*** | −0.010342*** | −0.010344*** |
| | (0.00317) | (0.00318) | (0.00313) | (0.00308) | (0.0031) | (0.00312) | (0.00309) | (0.00309) | (0.00312) | (0.00314) |
| $WB^2$ × IR | −0.000023 | −0.000026 | −0.000025 | −0.000028 | −0.000028 | −0.000027 | −0.000028 | −0.000028 | −0.000029 | −0.000029 |
| | (0.00004) | (0.00004) | (0.00004) | (0.00004) | (0.00004) | (0.00004) | (0.00004) | (0.00004) | (0.00004) | (0.00004) |
| SEAS_LEN | −0.00289*** | −0.002903*** | −0.002905*** | −0.002923*** | −0.002919*** | −0.002919*** | −0.00292*** | −0.002917*** | −0.002916*** | −0.002914*** |
| | (0.0002) | (0.0002) | (0.0002) | (0.0002) | (0.0002) | (0.0002) | (0.0002) | (0.0002) | (0.0002) | (0.0002) |
| ln(AREA) | −0.026715*** | −0.028156*** | −0.028593*** | −0.029355*** | −0.029372*** | −0.029907*** | −0.029856*** | −0.03008*** | −0.030234*** | −0.030394*** |
| | (0.00421) | (0.00418) | (0.00414) | (0.00405) | (0.00409) | (0.00407) | (0.00413) | (0.00411) | (0.00412) | (0.00416) |
| IR | 0.367149*** | 0.360023*** | 0.354778*** | 0.359568*** | 0.367855*** | 0.37482*** | 0.379161*** | 0.380267*** | 0.380272*** | 0.382034*** |
| | (0.05817) | (0.0586) | (0.05879) | (0.05892) | (0.05891) | (0.05883) | (0.05876) | (0.05823) | (0.05847) | (0.05852) |
| LON | 0.832721*** | 0.836214*** | 0.848977*** | 0.845548*** | 0.843861*** | 0.834363*** | 0.826556*** | 0.829392*** | 0.829003*** | 0.825721*** |
| | (0.16127) | (0.16081) | (0.16037) | (0.1611) | (0.16121) | (0.16086) | (0.16075) | (0.15972) | (0.1595) | (0.16012) |
| LAT | −0.273096 | −0.277614 | −0.279722 | −0.276367 | −0.27468 | −0.275366 | −0.276417 | −0.271148 | −0.266164 | −0.268191 |
| | (0.28759) | (0.28871) | (0.29112) | (0.30501) | (0.30837) | (0.31434) | (0.31533) | (0.32312) | (0.3275) | (0.32711) |
| PC_AREA_IR | 0.135243*** | 0.16083*** | 0.199129*** | 0.214396*** | 0.208581*** | 0.204158*** | 0.170041*** | 0.169686*** | 0.172422*** | 0.141826*** |
| | (0.01757) | (0.01957) | (0.01758) | (0.02002) | (0.02318) | (0.02167) | (0.02584) | (0.0239) | (0.02971) | (0.02905) |
| PC_AREA_FARM | 0.016772** | 0.011618* | 0.006148 | −0.004173 | −0.00166 | −0.008146 | −0.005238 | −0.015092 | −0.021428 | −0.024748 |
| | (0.00686) | (0.00684) | (0.00857) | (0.00974) | (0.00909) | (0.00998) | (0.01192) | (0.01394) | (0.0153) | (0.01755) |

*Note:* ***, ** and * indicate 1, 5 and 10 per cent significance levels, respectively. Robust standard errors are in parentheses, clustered at field level. Farm and crop fixed effects are included but not reported.Sample: 64,862 fields.